Description

Methods for Acquiring Shapes from HEp-2 Cell Sections and the Case-Based Recognition of HEp-2 Cells

The invention relates to methods for acquisition of shapes of images with representations of HEp-2 cell sections as objects and for learning abstract shape models from representations of HEp-2 cell sections for a case database for a case-based recognition of HEp-2 cells in the digital images; methods for acquisition of shapes of images with representations of HEp-2 cell sections as cases and for a case-based recognition of HEp-2 cells as objects in digital images; computer program products with a program code for performing these methods; computer program products on machine-readable carriers for performing these methods; and digital storage media that can interact with a programmable computer system in such a way such that these methods are carried out.

Arrangements for automatic examination of cells, cell complexes and other biological samples are disclosed, inter alia, in DE 196 16 997 A1 (method for automated microscope-supported examination of tissue samples or body fluid samples), DE 42 11 904 A1 (method and devices for providing a species list for a liquid sample), and DE 196 39 884 A1 (pattern recognition system).

According to DE 19616 997 A1, tissue samples or body fluid samples are examined with regard to cell types by application of a neuronal network.

Minute living beings such as worms, insects or snails are acquired and identified in DE 42 11 904 A1. The identification is done by comparison with objects contained in a reference object memory. At the same time, the identified objects are counted and inserted into a species list.

In DE 196 39 884 A1 solid components in a sample flow are acquired with regard

to their size in particular in accordance with their projection length in the image along the X axis and the Y axis, their circumference, and their average color density. The diagnostics by means of immunofluorescence according to the principle of fluorescence-optical assay of autoantibody binding is performed on frozen sections of HEp-2 cells. This method provides the most reliable results and provides a safe basis for therapeutic decisions.

A disadvantage is the currently missing automation so that a high personnel expenditure is required in connection with a health-hazardous, time-consuming evaluation that requires much experience.

An automated method is disclosed in DE 198 01 400 C2 (method and arrangement for automated recognition, property description, and interpretation of HEp-2 cell patterns). In this connection, only the shapes in the images are recognized. Automated inference in regard to other cases is not provided.

The invention specified in claims 1, 4 and 17 to 19 is based on the object of obtaining abstract shape models from images representing HEp-2 cell sections for a case database as well as of being able to determine objects automatically from digital images with objects by comparison with cases.

This object is solved by means of the features listed in the claims 1, 4 and 17-19.

The method for acquisition of shapes from images containing representations of HEp-2 cell sections with HEp-2 cells as objects and for learning abstract shape models of HEp-2 cells for a case database for the purpose of a case-based recognition of HEp-2 cells in digital images are distinguished in particular in that semi-automatically individual shapes of HEp-2 cell sections as objects are acquired in the form of representations in images and in that automatically, based on these individual shapes, abstract shape models in different abstract levels can be determined. The learned abstract shape models are either averaged shapes of

groups of objects or medians as individual shapes of groups of HEp-2 cells. The median is the object from which all other objects have the smallest spacing. The median is thereof a natural shape of an HEp-2 cell while the averaged shape is an artificial object that is not actually present.

The special advantage resides in that the contour or the shape is digitally detected and is saved in a datafile. Based on the data, subsequently manipulations can be carried out wherein, for example, similarity parameters can be determined and the similarity becomes describable.

Accordingly, this method is suitable for creating case databases with shape models of HEp-2 cells. Advantageously, groups of shapes of HEp-2 cells can be automatically generated and the similarities relative to one another can be hierarchically determined. From these groups, models of different abstraction levels can be determined moreover.

The basis are digital images with representations of HEp-2 cell sections with different appearance shapes with regard to contour and/or texture. By manually tracing edges that form contours and/or textures of an image by means of a hand-held input device connected to a computer, data are acquired that can be correlated with these edges and the thus represented HEp-2 cells as objects. Based on these data, shape models can be obtained in order to acquire knowledge in regard to the objects. In this way, the case database can be expanded advantageously.

In this connection, at least two HEp-2 cells are compared with one another, respectively, and are oriented toward one another and scaling and/or rotation is carried out.

Advantageously, at the same time the similarity is calculated wherein similarity parameters are acquired either as distance values or similarity values between the objects until either a minimum of the distance values or a maximum of the similarity

values is present.

Groups with similar shape models of HEp-2 cells can be learned or similar groups of HEp-2 cells can be combined wherein similarity relations can be generated as a comparison between these groups.

A further advantage resides in that with the application of the method, new shape models of HEp-2 cells of HEp-2 cell sections as objects in digital images can be also continuously correlated with the case database. In this way, an expansion of the case database is provided.

In this way, it is possible to generate a case database for automatic recognition, property description, and interpretation of HEp-2 cells in HEp-2 cell sections that serve for detection of autoimmune diseases. Autoimmune diseases are diseases that are characterized by a reactivity of the immune system against the body's own substances and structures. A common phenomenon in the case of autoimmune diseases is the occurrence of autoantibodies. The latter are immunoglobulins that are targeting the body's own structures. In addition to organ-specific autoantibodies, in particular, antibodies that are not organ-specific with reactivity against cellular structures are of importance. The detection of such autoantibodies is of great diagnostic importance.

For characterizing the specificity of autoantibodies, it is examined against which target antigens they are directed. This is possible with several methods. One of them is diagnostics by means of immunofluorescence. The latter is carried out on HEp-2 cells wherein the most reliable results are obtained. At the same time, it represents a safe basis for therapeutic decisions.

The case database for automatic recognition, property description, and interpretation of HEp-2 cells in HEp-2 cell sections are used for the detection of autoimmune diseases. Autoimmune diseases are diseases that are characterized

by a reactivity of the immune system against the body's own substances and structures. A common phenomenon in the case of autoimmune diseases is the occurrence of autoantibodies. The latter are immunoglobulins that are targeting the body's own structures. In addition to organ-specific autoantibodies, in particular, antibodies that are not organ-specific with reactivity against cellular structures are of importance. The detection of such autoantibodies is of great diagnostic importance.

For characterizing the specificity of autoantibodies, it is examined against which target antigens the autoantibodies are directed. This is possible with several methods. One of them is diagnostics by means of immunofluorescence. The latter is carried out on HEp-2 cells wherein the most reliable results are obtained. At the same time, it represents a safe basis for therapeutic decisions.

The case database thus is the basis for automated case-based recognition and determination of HEp-2 cells in HEp-2 cell sections as objects in digital images with objects.

The selected case image and the generated gradient image of the digital image with objects is transformed into pyramids with image planes. The individual image planes are compared sequentially with one another wherein the highest image planes are used first. The highest image planes are the least sharp image planes with the least amount of data, respectively, so that the comparison is carried out beginning with the least computing expenditure. Furthermore, the selected case image is compared successively with every object of the digital image with objects. During the comparison between each object image and each case image, an orientation and scaling and/or rotation of the case image is carried out wherein at same time the similarity is calculated.

The special advantage resides in that either the contour or the shape can be digitally detected and saved in a data file. With this data, subsequently

manipulations can be carried out wherein, for example, similarity parameters can be determined and the similarity and/or the similarity as a degree of matching between the case image and object image are describable by the similarity parameter. With decreasing similarity parameter, the object image becomes less similar to the case image.

The methods according to the invention can be advantageously made available to the users as computer program products with a program code, respectively, for performing this method, as computer program products on machine-readable carriers for performing these methods, and as digital storage media that can cooperate with a programmable computer system.

Advantageous embodiments of the invention are set forth in claims 2, 3, and 5 to 16.

Advantageously, according to the embodiment of claim 2, the distance values and/or similarity values provide a distance matrix or similarity matrix.

In accordance with the embodiment of claim 3, the distance values or the similarity values are advantageously hierarchically represented by means of single linkage method and a dendogram.

According to the embodiment of claim 5, at least two cases are compared with one another, respectively, wherein the cases are oriented toward one another and scaling and/or rotation is carried out. Advantageously, at the same time the similarity is calculated, wherein similarity parameters are determined either as distance values or similarity values between the cases, respectively, until either a minimum of the distance values or a maximum of the similarity values is present.

Advantageously, according to the embodiment of claim 6, the dendogram on the similarity scale is intersected at least once according to either determined, and thus

automatic, or user-specific thresholds so that groups will be generated. The individual shapes are correlated with the groups and within the groups a prototype is selected, wherein the prototype is either an averaged shape that is averaged based on the individual shapes of the group or the median of the group of individual shapes. In this way, a visual control of individual groups and/or the individual objects is provided. The averaged shape or the median of the group is represented on the data viewing device and its contour points are saved as a data set in the computer.

In accordance with the embodiment of claim 7, advantageously a reduction of the data obtained by tracing the edges and thus of the points as the visible outer and/or inner contours is done by interrelation with a polynomial.

The cases correlated with the scanned edges are advantageously transformed in accordance with the embodiment of claim 8 such that the center point of a case corresponds to the coordinate origin 0, 0. The cases are aligned within a coordinate system so that a comparison with regard to their similarities relative to one another is possible easily.

The calculation of the similarities is based on the determination of similarity parameters. In this connection, at least one case and one object are compared with one another wherein they are oriented toward one another and scaling and/or rotation is carried out. At the same time, the similarity is calculated wherein, according to the embodiment of claim 9, similarity parameters are determined either as distance values or similarity values between the case and the object, respectively, until either a minimum of the distance values or a maximum of the similarity values is present.

Advantageously, in accordance with the embodiment of claim 10, a gradient image is generated by means of edge detection of the objects of the digital image wherein gradients are assigned to large changes of the grayscale in vertical direction as well

as in horizontal direction while no gradients are assigned to homogenous surfaces. The homogenous surfaces are black. The result is an image with white edges of the object while the surfaces enclosed by the edges of the objects and the surfaces adjoining the edges of the object are black. The data set of the digital image is thus significantly reduced in comparison to a grayscale image of the digital image. At the same time, the computation expenditure is reduced for comparing each object with the selected case by the calculation of the similarities with the determination of the similarity parameters. Moreover, stacked and partially overlapping objects in the digital image can be more easily determined by comparison with a selected case.

According to the embodiment of claim 11, a gradient image is generated based on the case image as well as the object image, respectively, and the gradient images each are transformed into an image sequence as a pyramid with image planes, wherein successively the directional vectors in the image planes of the case image and the object image, respectively, are compared with one another by forming the product. The principle of pyramids reduces the computational expenditure significantly. The respective subsequent image planes of the pyramids are representations of a raster that is twice as coarse. For this purpose, only every other point of a line and only every other line are picked and combined to a new image as an image plane. The employed tracing theorem ensures at the same time that the original finer raster can be exactly reconstructed based on the coarser raster. When comparing the cases and the object, one starts advantageously with the coarsest raster of the uppermost image planes. Depending on the result of the comparison of the similarity, successively image planes are compared with one another with the finer raster, respectively. The comparison can be interrupted at any time so that the computational expenditure for the comparison can be reduced significantly.

According to the embodiment of claim 12, a dendogram represents an advantageous differentiation means of individual cases wherein groups of individual cases are hierarchically ordered. In this connection, the case image is a prototype

of a group of individual cases wherein the groups are sets of similar individual cases with certain distance values or similarity values. The most similar case determines the branch of the dendogram with similar cases for determination of the object. The prototype is either an averaged shape averaged from the individual shapes of the group or the median of the group of individual shapes. The median is the case from which all other cases have the smallest spacing. The median represents thus a natural HEp-2 cell while the averaged shape is an artificial HEp-2 cell. However, the case image can also be an individual image of an object.

By means of determining the directional vector between either two points or neighboring points of the edges either in the case image or in the object image in accordance with the embodiment of claim 13, the direction of the edge as a local orientation is determined so that the components of the course of direction of the local orientation and of the expression, measured e.g. based on the height or the slant of the edge, are incorporated into description of the edge. In this way, advantageously also the surroundings of the existing complex structure of the image information can be incorporated into the calculation of the similarities. In the calculation of similarities, the similarity parameters are thus determined as directional vectors as well as distance values or similarity values between the case image and the object image. Additional information of the digital image is thus advantageously also considered when comparing by means of calculation of similarity.

According to the further embodiment of claim 14, either the prototypes or the cases are ordered by means of an index in accordance with the similarity relations in the case database. The index characterizes an index register with the prototypes and/or the cases individually or in groups wherein from a set of prototypes and/or cases the most similar prototype or case in comparison to the object in the image can be found quickly.

The calculation of similarity is realized advantageously according to the formula of

the embodiment of claim 15.

The embodiment according to claim 16 advantageously provides that a non-identical object can be determined as a case manually and assigned to the dendogram with the determined cases. In this way, the case database can be expanded continuously.

Embodiments of the invention will be explained in more detail in the following with the aid of the illustrations.

It is shown in:

Fig. 1 an illustration with labeled and approximated contours of sectioned HEp-2 cells;

Fig. 2 the illustration with numbered representations of the sectioned HEp-2 cells of Fig. 1; and

Fig. 3 a dendogram of these HEp-2 cells.

A method for acquisition of shapes of digital images of HEp-2 cell sections with HEp-2 cells as objects and for learning abstract shape models of HEp-2 cells for a case database for case-based recognition of HEp-2 cells in digital images will be explained in more detail as an example of the invention in a first embodiment. In this context, HEp stands for human epithelium.

On a data viewing device in the form of a known monitor connected to a computing device such as a computer a digital image of an HEp-2 cell section is represented whose contours can be different.

Fig. 1 shows an illustration with labeled and approximated contours of sectioned HEp-2 cells.

By manual tracing of edges of the digital image with a handheld input device in

connection with the data viewing device, data is acquired that can be correlated with the scanned edges. Edges are in this connection visible outer and/or inner contours of represented sectioned HEp-2 cells as objects.

By means of the handheld input devices in the form of the cursor of the monitor guided by a keyboard or a mouse, a light pen with a photodetector, a scanner and/or a pen and scanner in cooperation with the data viewing device, data of the contour points as X-coordinate, Y-coordinate and/or grayscale values or color values of the contour points are acquired as data of the edges of HEp-2 cells assignable to the objects. A further embodiment is provided by a combination of a pen in connection with a touchscreen. Such input devices and the method for obtaining the data correlated with the edges are known so that no further explanation is needed.

Tracing of the contours is done manually by means of the handheld input device on a digital image represented on a data viewing device. The traced contours can also be represented with labels by means of the data viewing device wherein this is represented as an area field at least partially by the scanned contour and/or edge on the data viewing device. In this way, easy control of the traced edges is possible on the data viewing device. Errors that are caused, for example, by lack of concentration, disruptions, distractions, or fatigue of the persons tracing the contours and/or edges of the cases are prevented.

Each of the HEp-2 cells as objects determined by the edges is scaled in a coordinate system wherein the center point of the object corresponds to the coordinate origin X = 0 and Y = 0.

The similarity of sectioned HEp-2 cells is determined by paired orientation relative to one another, respectively, until the similarity parameter no longer changes. Scaling and/or rotation is performed wherein at the same time the similarities are calculated. During calculation of the similarity the similarity parameters are

determined either as distance values or as similarity values between the objects, respectively, until either a minimum of the distance values or a maximum of the similarity values is present. The calculation of similarity is done in accordance with

$$D(P,O) = \sum_{i=1}^{N} \left| \frac{(p_i - \mu_p)}{\delta_p} - R(\Theta) \frac{(o_i - \mu_o)}{\delta_o} \right|^2$$

wherein

P and O - the objects

Θ - the rotation matrix,

μ_p and μ_o - the center points of the objects P and O, and

δ_p and δ_o - the sums of the squared spacings of each point from the center points.

The distance values or the similarity values define a distance matrix or a similarity matrix.

Based on the determined similarity values, sets of similar objects are formed and are ordered hierarchically as a dendogram. Fig. 2 shows numbered illustrations of the sectioned HEp-2 cells of Fig. 1. The dendogram is intersected with the similarity scale in accordance either with fixed or user-specific thresholds so that groups will result. In the case of fixed thresholds, the dendogram is intersected automatically. The individual shapes are assigned to the groups and a prototype is selected within the groups, respectively. The prototype is either an averaged shape which is averaged from the individual shapes of the group or the median of the group of the individual shapes. The prototype of the group is represented on the data viewing device and the contour points of the prototypes are saved as data sets in the computer. Fig. 3 shows a dendogram of this HEp-2 cell section.

This method is employed on further digital images so that a case database with shape models as prototypes with averaged shapes of groups of individual shapes and/or with mediums of groups of individual shapes will result.

In one variant of the embodiment the data of edges of visible outer and/or inner contours in the digital image, acquired by means of the input device in connection with the data viewing device, are reduced by interpolation. According to this interpolation:

- in a first step, the starting point is assigned to a first point of the edge of the object which first point is determined by the coordinate system and therefore scaled;
- in a second step, a virtual line to a neighboring point as the second point is drawn;
- in a third step, the distance between this virtual line and the corresponding segment of the contour of a preceding object is determined,
- in a fourth step, this spacing is compared as a value with a preset value; and
- in a fifth step, a starting point for a virtual line to the next point is assigned to the second point.

The steps 3, 4, and 5 are repeated for the entire contour of the object.

The distance values or similarity values can be advantageously hierarchically represented in accordance with a further embodiment by means of single linkage method and a dendogram.

A method for acquisition of shapes of digital images of HEp-2 cells sections with HEp-2 cells as objects and for learning abstract shape models of HEp-2 cells will be explained in more detail as one example of the invention in a second embodiment.

The case database with cases as case images, determined by a method for

acquisition of shapes of digital images of HEp-2-cell sections with HEp-2 cells as objects and for learning abstract shape models of HEp-2 cells of the first embodiment, form the basis for a case-based recognition of HEp-2 cells as objects in digital images with objects.

A case image with a case description is selected from the case database. The case image is either a prototype of a group of individual cases or an individual image of a case. The group of individual cases represent sets of similar individual cases with certain distance values or similarity values that are ordered hierarchically as a dendogram. The most similar case determines the branch of the dendogram. The prototype itself is either an averaged shape determined based on the individual shapes of the group or the median of the group of individual shapes. The case image is transformed into an image sequence as a pyramid with image planes of the case image. An image sequence as a pyramid with image planes prevents an explosive growth of the computation expenditure. The case image can be represented sequentially on a raster that is twice as coarse without any loss of information by application of smoothing operations, wherein all wave numbers remain below half the limit wave number, and based on the tracing theorem. In this connection only every other point of the line and only every other line are picked and combined to a new image wherein it is ensured that the original finer raster can be exactly reconstructed from the coarser raster. The application of the smoothing operations is done iteratively so that based thereon a sequence of images results wherein the images with regard to their surface area become smaller by the factor 4. The image planes that become smaller and smaller result in the shape of a pyramid when stacked.

Based on the actual digital image with sectioned HEp-2 cells as objects a gradient image is generated. By means of edge detection of the objects of the digital image comprising objects the gradient image is generated wherein gradients will be assigned to large changes of the grayscale value in the vertical direction as well as horizontal direction and no gradients will be assigned to homogenous surfaces. The

homogenous surfaces are thus black. The gradient image is transformed also into an image sequence as a pyramid with image planes.

The case image is subsequently successively moved onto each object image of the gradient image beginning with the highest image plane of the case image and the object image wherein the case image is compared with each object image of the gradient image. During comparison, the case image is oriented toward the object image wherein in this connection scaling and/or rotation of the case image is done. When comparing the case image with the object image the similarity between the case image and object image is calculated at the same time. In the calculation of similarity, the similarity parameters are determined either as distance values or similarity values between the case image and the object image, respectively, until either a minimum of the distance values or a maximum of the similarity values is present. The similarity parameters determine the degree of similarity between the case image and object image wherein the degree of similarity decreases with decreasing similarity parameter and the object image becomes less similar to the case image.

In a further embodiment, the directional vector between either two points or neighboring points of the edges either has been calculated for the case image or will be calculated for the object image. When calculating the similarity, the similarity parameters as well as the directional vectors as well as either distance values or similarity values between the case image and the object image are determined.

The HEp-2 cells as cases are ordered by means of an index in accordance with the similarity relations in such a way in the case database that either the most similar prototype from a set of prototypes or the most similar case from a set of cases can be found quickly for the object in the image.

The prototype either as an averaged shape or median of the group or the individual image is represented on a data viewing device in the form of a monitor connected

to a computer in which the method is performed. Moreover, the contour points either of the averaged shape or the median or the individual image are saved as a data set in the computer.

In a further variant of the embodiment a gradient image is formed from the case image and the object image, respectively. These gradient images are transformed, respectively, into an image sequence as a pyramid with image planes and, successively, the directional vectors in the image plans of the case image and the object image, respectively, are compared with one another by forming the product.

A third embodiment is a computer program product with a program code for performing

- either a method disclosed in the first embodiment for acquisition of shapes from digital images with representations of HEp-2 cell sections and for learning abstract shape models of the represented HEp-2 cells
- or a method disclosed in the second embodiment for acquisition of shapes of images with representations of HEp-2 cell sections as cases and for case-based recognition of HEp-2 cells as objects in the digital images,

when the program is run on a computer.

The fourth embodiment is a computer program product on the machine-readable carrier for performing

- either a method disclosed in the first embodiment for acquisition of shapes from digital images with representations of HEp-2 cells and for learning abstract shape models from representations of HEp-2 cell sections
- or a method disclosed in the second embodiment for acquisition of shapes of images with representations of HEp-2 cell sections as cases and for case-based recognition of HEp-2 cells as objects in the digital images,

when the program is run on a computer.

A fifth embodiment is a digital storage medium that interacts with a programmable

computer system such that

- either a method disclosed in the first embodiment for acquisition of shapes from digital images with representations of HEp-2 cells and for learning abstract shape models from representations of HEp-2 cell sections
- or a method disclosed in the second embodiment for acquisition of shapes from images with representations of HEp-2 cell sections as cases and for case-based recognition of HEp-2 cells as objects in the digital images,

is carried out.